Date: 9/19/01 Express Mail Label No. EL552577526US

Inventors: Sridhar Ramaswamy, Todd R. Golub, Pablo Tamayo and Michael Angelo

Attorney's Docket No.: 2825.2020-002

GENETIC MARKERS FOR TUMORS

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Nos. 60/233,534, filed on September 19, 2000, and 60/278,749, filed on March 26, 2001. The entire teachings of the above applications are incorporated herein by reference.

GOVERNMENT SUPPORT

The invention was supported, in whole or in part, by grant NIH-5T32HL07623 from the National Institutes of Health. The U.S. Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

Classification of tumor samples from individuals is not an exact science. In many instances, accurate diagnosis and safe and effective treatment of a disorder depends on being able to discern biological distinctions among morphologically similar samples, such as tumor samples. The classification of a sample from an individual into particular disease classes has typically been difficult and often incorrect or inconclusive. Using traditional methods, such as morphology analyses, histochemical analyses, immunophenotyping and cytogenetic analyses, often only one or two characteristics of the sample are analyzed to determine the sample's classification, resulting in inconsistent and sometimes inaccurate results. Such results can lead to incorrect

diagnoses and potentially ineffective or harmful treatment. Thus, a need exists for accurate markers for identifying tumor classes and classifying tumor samples.

SUMMARY OF THE INVENTION

As described herein, sets of genetic markers which are specific to various tumor classes have been identified. The patterns of expression for these genes will be useful in improving the diagnosis and classification of human cancer. This information will be useful for designing genetic or antibody-based tests for the characterization of clinical tumor samples, and in particular, those samples that are difficult to evaluate with present histopathologic techniques. In addition, a number of specific markers may encode secreted or membrane bound proteins. These proteins would prove useful for the early detection of cancer (analogous to the serum prostate specific antigen (PSA) test) or for the treatment of cancer (analogous to antibody-based treatment of breast cancer by targeting the Her-2/Neu gene product). Finally, genes which are specifically expressed by classes of cancer may be involved in the pathogenesis of disease and are potential therapeutic targets.

The invention relates to classification or identification of biological samples, e.g., tumor samples, based on the simultaneous expression monitoring of a set of genes as described herein using DNA microarrays or other methods developed to assess a large number of genes. Microarrays have the attractive property of allowing one to monitor multiple expression events in parallel using a single technique. The method can be used to distinguish among tumor samples (e.g., to distinguish a breast tumor sample from a prostate tumor sample) or between a tumor sample and corresponding normal sample (e.g., to distinguish a breast tumor sample from a normal breast tissue sample) based on the patterns of gene expression of the samples. The markers identified herein can also be used to classify or identify tumors of unknown primary origin. The invention also relates to classification or identification of biological samples, e.g., tumor samples, based on the expression of a set of proteins encoded by a set of marker genes as described herein.

Both nucleic acid- and protein-based monitoring methods of the genes identified in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3 (or their encoded proteins) can be used to predict or aid in the prediction of, diagnose or aid in the diagnosis of, or monitor or aid in the monitoring of cancer, particularly tumor, establishment, progression or regression in an individual.

In one aspect, the invention features a method of identifying a tumor comprising the steps of: a) obtaining a sample derived from an organ or tissue; b) determining the expression pattern of one or more marker genes in the sample, said one or more marker genes selected from the group consisting of the genes in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3; and c) comparing the expression pattern obtained in step b) to the expression pattern of one or more genes specific to a tumor. A marker gene expression pattern in the sample that is similar to the gene expression pattern specific to a tumor identifies a tumor.

In one embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 1A-1R2, whereby the tumor identified is a bladder tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 2A-2T2, whereby the tumor identified is a breast tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 3A-3Z2, whereby the tumor identified is a central nervous system (CNS) tumor. In yet another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 4A-4S2, whereby the tumor identified is a colorectal tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 5A-5M2, whereby the tumor identified is leukemia. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 6A-

6W2, whereby the tumor identified is a lung tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 7A-7D3, whereby the tumor identified is a lymphoma. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 8A-8X2, whereby the tumor identified is a melanoma. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 9A-9C3, whereby the tumor identified is a mesothelioma. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 10A-10P2, whereby the tumor identified is an ovarian tumor. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in 11A-11O2, whereby the tumor identified is a pancreatic tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 12A-12V2, whereby the tumor identified is a prostate tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 13A-13N2, whereby the tumor identified is a renal tumor. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 14A-14A3, whereby the tumor identified is a uterine tumor.

In other embodiments, the marker gene is DNA or it corresponding mRNA. Preferably, when the marker gene is DNA or mRNA, the gene expression pattern of the marker gene is determined utilizing specific hybridization probes. For example, the gene expression pattern may be determined utilizing oligonucleotide microarrays.

In another embodiment, the marker genes are expressed as polypeptides. Preferably, when the marker genes are expressed as polypeptides, the gene expression pattern is determined utilizing antibodies.

In another aspect, the invention features a method of predicting the likelihood of tumor development in a subject, comprising the steps of: a) obtaining a sample derived from an organ or tissue of a subject; b) determining the expression pattern of one or more marker genes in the sample, said one or more marker genes selected from the

group consisting of the genes in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3; and c) comparing the expression pattern obtained in step b) to the expression pattern of one or more genes specific to a tumor. A marker gene expression pattern in the sample that is similar to the gene expression pattern specific to a tumor indicates an increased likelihood of tumor development in the subject.

In one embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 1A-1R2, whereby the tumor for which a likelihood of development is predicted is a bladder tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 2A-2T2, whereby the tumor for which a likelihood of development is predicted is a breast tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 3A-3Z2, whereby the tumor for which a likelihood of development is predicted is a central nervous system (CNS) tumor. In yet another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 4A-4S2, whereby the tumor for which a likelihood of development is predicted is a colorectal tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 5A-5M2, whereby the tumor for which a likelihood of development is predicted is leukemia. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 6A-6W2, whereby the tumor for which a likelihood of development is predicted is a lung tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 7A-7D3, whereby the tumor for which a likelihood of development is predicted is a lymphoma. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 8A-8X2, whereby the tumor for which a likelihood of development is predicted is a melanoma. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 9A-9C3, whereby the tumor for

which a likelihood of development is predicted is a mesothelioma. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 10A-10P2, whereby the tumor for which a likelihood of development is predicted is an ovarian tumor. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in 11A-11O2, whereby the tumor for which a likelihood of development is predicted is a pancreatic tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 12A-12V2, whereby the tumor for which a likelihood of development is predicted is a prostate tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 13A-13N2, whereby the tumor for which a likelihood of development is predicted is a renal tumor. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 14A-14A3, whereby the tumor for which a likelihood of development is predicted is a uterine tumor.

In other embodiments, the marker gene is DNA or it corresponding mRNA. Preferably, when the marker gene is DNA or mRNA, the gene expression pattern of the marker gene is determined utilizing specific hybridization probes. For example, the gene expression pattern may be determined utilizing oligonucleotide microarrays.

In another embodiment, the marker genes are expressed as polypeptides. Preferably, when the marker genes are expressed as polypeptides, the gene expression pattern is determined utilizing antibodies.

In still another aspect, the invention features a method of diagnosing a tumor in a subject, comprising the steps of: a) obtaining a sample derived from an organ or tissue of a subject; b) determining the expression pattern of one or more marker genes in the sample, said one or more marker genes selected from the group consisting of the genes in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A32; and c) comparing the expression pattern obtained in step b) to the expression pattern of one or more genes

specific to a tumor. A marker gene expression pattern in the sample that is similar to the gene expression pattern specific to a tumor indicates the presence of a tumor in the subject.

In one embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 1A-1R2, whereby the tumor that is diagnosed is a bladder tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 2A-2T2, whereby the tumor that is diagnosed is a breast tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 3A-3Z2, whereby the tumor that is diagnosed is a central nervous system (CNS) tumor. In yet another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 4A-4S2, whereby the tumor that is diagnosed is a colorectal tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 5A-5M2, whereby the tumor that is diagnosed is leukemia. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 6A-6W2, whereby the tumor that is diagnosed is a lung tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 7A-7D3, whereby the tumor that is diagnosed is a lymphoma. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 8A-8X2, whereby the tumor that is diagnosed is a melanoma. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 9A-9C3, whereby the tumor that is diagnosed is a mesothelioma. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 10A-10P2, whereby the tumor that is diagnosed is an ovarian tumor. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in 11A-11O2, whereby the tumor that is diagnosed is a pancreatic tumor. In another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 12A-12V2, whereby the tumor that is diagnosed is a prostate tumor. In another embodiment,

the one or more marker genes are selected from the group consisting of the genes in FIGS. 13A-13N2, whereby the tumor that is diagnosed is a renal tumor. In still another embodiment, the one or more marker genes are selected from the group consisting of the genes in FIGS. 14A-14A3, whereby the tumor that is diagnosed is a uterine tumor.

In other embodiments, the marker gene is DNA or it corresponding mRNA. Preferably, when the marker gene is DNA or mRNA, the gene expression pattern of the marker gene is determined utilizing specific hybridization probes. For example, the gene expression pattern may be determined utilizing oligonucleotide microarrays.

In another embodiment, the marker genes are expressed as polypeptides. Preferably, when the marker genes are expressed as polypeptides, the gene expression pattern is determined utilizing antibodies.

In yet another aspect, the invention features a method of identifying a compound for use in treating cancer, comprising the steps of: a) providing a cell or cell lysate sample; b) contacting the cell or cell lysate sample with a candidate compound; and c) detecting a decrease in expression of one or more genes specific to a tumor, said one or more genes selected from the group consisting of the genes in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3. A candidate compound that decreases the expression of one or more genes specific to a tumor identifies a compound for use in treating cancer.

In one embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 1A-1R2, whereby the compound identified is useful for treating bladder cancer. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 2A-2T2, whereby the compound identified is useful for treating breast cancer. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 3A-3Z2, whereby the compound identified is useful for treating central nervous system (CNS) cancer. In yet another embodiment, the one or more genes are selected from the group consisting of

the genes in FIGS. 4A-4S2, whereby the compound identified is useful for treating colorectal cancer. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 5A-5M2, whereby the compound identified is useful for treating leukemia. In still another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 6A-6W2, whereby the compound identified is useful for treating lung cancer. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 7A-7D3, whereby the compound identified is useful for treating lymphoma. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 8A-8X2, whereby the compound identified is useful for treating melanoma. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 9A-9C3, whereby the compound identified is useful for treating mesothelioma. In still another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 10A-10P2, whereby the compound identified is useful for treating ovarian cancer. In still another embodiment, the one or more genes are selected from the group consisting of the genes in 11A-11O2, whereby the compound identified is useful for treating pancreatic cancer. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 12A-12V2, whereby the compound identified is useful for treating prostate cancer. In another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 13A-13N2, whereby the compound identified is useful for treating renal cancer. In still another embodiment, the one or more genes are selected from the group consisting of the genes in FIGS. 14A-14A3, whereby the compound identified is useful for treating uterine cancer.

In other embodiments, the gene is DNA or it corresponding mRNA. Preferably, when the marker gene is DNA or mRNA, the gene expression pattern of the marker gene is determined utilizing specific hybridization probes. For example, the gene expression pattern may be determined utilizing oligonucleotide microarrays.

In another embodiment, the genes are expressed as polypeptides. Preferably, when the marker genes are expressed as polypeptides, the gene expression pattern is determined utilizing antibodies.

In another aspect, the invention features an oligonucleotide microarray having immobilized thereon a plurality of oligonucleotide probes specific for one or more tumor specific genes selected from the group consisting of the genes in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3.

In preferred embodiments, the oligonucleotide probes specific for one or more tumor specific genes are selected from the genes in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3, respectively.

In other embodiments, the oligonucleotide probes are DNA or mRNA.

The invention also features a method for modulating tumor development in a subject by decreasing in the subject at least one marker gene shown to be specific to a particular tumor class, for example, any of the marker genes shown herein.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1R2 are a table of marker genes for bladder tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (bladder) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the

corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS 2A-2T2 are a table of marker genes for breast tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (breast) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 3A-3Z2 are a table of marker genes for central nervous system (CNS) tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (CNS) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 4A-4S2 are a table of marker genes for colorectal tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (colorectal) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth

columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 5A-5M2 are a table of marker genes for leukemia. The second column of the table (entitled "Distinction") shows the type of tumor (leukemia) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 6A-6W2 are a table of marker genes for lung tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (lung) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 7A-7D3 are a table of marker genes for lymphoma tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (lymphoma) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 8A-8X2 are a table of marker genes for melanoma tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (melanoma) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 9A-9C3 are a table of marker genes for mesothelioma tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (mesothelioma) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled

"Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 10A-10P2 are a table of marker genes for ovarian tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (ovarian) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 11A-11O2 are a table of marker genes for pancreatic tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (pancreatic) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 12A-12V2 are a table of marker genes for prostate tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (prostate) for which the marker gene is specific. The third column (entitled "Distance") shows the

signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 13A-13N2 are a table of marker genes for renal tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (renal) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 14A-14A3 are a table of marker genes for uterine tumor types. The second column of the table (entitled "Distinction") shows the type of tumor (uterine) for which the marker gene is specific. The third column (entitled "Distance") shows the signal-to-noise distance, which is an indication of the robustness of the marker; the larger the number, the more robust (specific) the marker. The fourth, fifth and sixth columns show the result of permutation tests which are indicators of the possibility that the marker would appear by chance. The seventh column (entitled "Feature") shows the designation assigned to that marker on the Affymetrix microarray used as described in the Examples. This designation corresponds to a GenBank Accession number for the

corresponding gene. The eighth column (entitled "Desc.") provides descriptive information about the marker gene.

FIGS. 15-27 each show gene order as a function of measure of correlation for a variety of tumors.

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to the identification of sets are marker genes which are specific for particular tumor classes. The marker genes for particular tumor types are shown in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3.

In one embodiment, the genetic markers described herein can be used to identify or classify tumors, such as tumors of unknown primary derivation. In this embodiment, a tumor sample is obtained and the gene expression pattern of a set of genes identified in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3 is determined. For example, the nucleic acid molecules within the sample can be rendered available for hybridization to an oligonucleotide array as described in the Examples. Alternatively, the expression of the proteins encoded by a set of marker genes identified herein can be assessed, e.g., using antibody-based methods. The marker genes (or encoded proteins) to be assessed can be all or a portion of the marker genes associated with a single particular tumor class, or can be all or a portion of the marker genes associated with several different tumor classes.

The expression pattern obtained can then be compared with the expression pattern(s) associated with one or more classes of tumors as described herein, and a classification of the tumor can be made based on the similarity or identity of the sample expression pattern and the pattern characteristic of a particular tumor class. For example, it may be determined that the expression pattern of the marker genes tested

correlates most closely with the expression pattern characteristic of tumors of the breast, and a determination can be made that the most likely primary derivation of the tumor sample is breast.

By "gene expression pattern" is meant the level or amount of gene expression of particular genes, for example, marker genes as assessed by methods described herein. The gene expression pattern can comprise data for one or more genes and can be measured at a single time point or over a period of time. For example, the gene expression pattern can be determined using a single marker gene, or it can be determined using two or more marker genes, three or more marker genes, five or more marker genes, eight or more marker genes, twenty or more marker genes, or fifty or more marker genes. A gene expression pattern may include expression levels of marker genes that are not specific to a particular tumor or tumor class, as well as genes that are specific to a particular tumor or tumor class. Classification (e.g., the presence or absence of tumor, or the identification of a compound that modulates tumor development) can be made by comparing the gene expression pattern of the sample with respect to one or more marker genes with one or more gene expression patterns specific to a particular tumor or tumor class (e.g., in a database). Using the methods described herein, expression of numerous genes can be measured simultaneously. The assessment of numerous genes provides for a more accurate evaluation of the sample because there are more genes that can assist in classifying the sample.

As used herein, "marker genes" are proteins, polypeptides, or nucleic acid molecules (e.g., mRNA, tRNA, rRNA, cDNA, or cRNA) that result from transcription or translation of genes. The present invention can be used effectively to analyze proteins, polypeptides, or nucleic acid molecules that are the result of transcription or translation, particularly of the genes identified herein. The nucleic acid molecule levels measured can be derived directly from the gene or, alternatively, from a corresponding regulatory gene or regulatory sequence element. All forms of marker genes can be measured. For example, the nucleic acid molecule can be transcribed to obtain an RNA gene expression product. If desired, the transcript can be translated using, for example,

standard *in vitro* translation methods to obtain a polypeptide gene expression product. Polypeptide marker gene products can be used in protein binding assays, for example, antibody assays, or in nucleic acid binding assays, standardly known in the art, in order to identify tumors or compounds involved in tumor development. Additionally, variants of marker genes including, for example, spliced variants and polymorphic alleles, can be measured. Similarly, gene expression can be measured by assessing the level of a polypeptide or protein or derivative thereof translated from mRNA. The sample to be assessed can be any sample that contains a marker gene. Suitable sources of marker genes, e.g., samples, can include intact cells, lysed cells, cellular material for determining gene expression, or material containing gene expression products. Examples of such samples are cells or tissue derived from the bladder, breast, CNS, colorectal, blood, bone marrow, lung, lymphatic system, skin, mesothelium, ovary, pancreas, prostate, kidney, or uterus. Methods of obtaining such samples are known in the art.

In one embodiment, the marker gene is a protein or polypeptide. As used herein, by "polypeptide" is meant any chain of more than two amino acids, regardless of post-translational modification such as glycosylation or phosphorylation. Examples of polypeptides include, but are not limited to, proteins. In this embodiment the determination of the gene expression pattern is made using techniques for protein detection and quantitation known in the art. For example, antibodies that specifically interact with the protein or polypeptide expression product of one or more genes specific to a particular tumor or tumor class can be obtained using methods that are routine in the art. The specific binding of such antibodies to protein or polypeptide gene expression products can be detected and measured by methods known in the art, for example, Western blot analysis or ELISA techniques.

In a preferred embodiment, the marker is a nucleic acid, for example, DNA or mRNA, and the gene expression levels are obtained by contacting the sample with a suitable microarray on which probes specific for all or a subset of the genes specific to a particular tumor or tumor class have been immobilized, and determining the extent of

hybridization of the nucleic acid in the sample to the probes on the microarray. Such microarrays are also within the scope of the invention. Examples of methods of making oligonucleotide microarrays are described, for example, in WO 95/11995. Other methods are readily known to the skilled artisan.

As used herein, "genes specific to a particular tumor or tumor class," refers to a gene or genes whose expression correlates with a particular type of tumor. Expression patterns obtained for genes specific to a particular tumor or tumor class can be used to determine, for example, the presence or absence of a particular tumor in a sample, or if a candidate compound increases or decreases gene expression in a sample. Samples can be classified according to their broad expression pattern, or according to the expression levels of particular genes specific to a particular tumor or tumor class. The genes that are relevant for classification are referred to herein as "genes specific to a particular tumor or tumor class." Not all genes specific to a particular tumor or tumor class for a particular class distinction must be assessed in order to classify a sample. A subset of the genes specific to a particular tumor or tumor class that demonstrate a high correlation with a tumor class distinction can be used in classifying the presence of an that particular tumor type. This subset can be, for example, one or more genes, two or more genes, three or more genes, five or more genes, eight or more genes, twenty or more genes, or fifty or more genes. The genes specific to a particular tumor or tumor class that characterize other classification categories such as, for example, a candidate compound that modulates tumor development, can be the same or different from the genes specific to a particular tumor or tumor class that characterize the presence or absence of a tumor. Typically the accuracy of the classification increases with the number of genes specific to a particular tumor or tumor class that are assessed.

The gene expression value measured or assessed is the numeric value obtained from an apparatus that can measure gene expression levels. Gene expression levels refer to the amount of expression of the gene expression product, as described herein. The values are raw values from the apparatus, or values that are optionally re-scaled, filtered and/or normalized. Such data is obtained, for example, from a GeneChip®

probe array or Microarray (Affymetrix, Inc.; U.S. Patent Nos. 5,631,734, 5,874,219, 5,861,242, 5,858,659, 5,856,174, 5,843,655, 5,837,832, 5,834,758, 5,770,722, 5,770,456, 5,733,729, 5,556,752, all of which are incorporated herein by reference in their entirety), and the expression levels are calculated with software (e.g., Affymetrix GENECHIP software). For example, nucleic acids (e.g., mRNA or DNA) from a sample that has been subjected to particular stringency conditions hybridize to the probes on the chip. The nucleic acid to be analyzed (e.g., the target) is isolated, amplified and labeled with a detectable label, (e.g., ^{32}P or fluorescent label) prior to hybridization to the arrays. After hybridization, the arrays are inserted into a scanner that can detect patterns of hybridization. These patterns are detected by detecting the labeled target now attached to the microarray, e.g., if the target is fluorescently labeled, the hybridization data are collected as light emitted from the labeled groups. Since labeled targets hybridize, under appropriate stringency conditions known to one of skill in the art, specifically to complementary oligonucleotides contained in the microarray, and since the sequence and position of each oligonucleotide in the array are known, the identity of the target nucleic acid applied to the probe is determined.

Quantitation of gene expression patterns from the hybridization of a labeled nucleic acid microarray can be performed by scanning the microarray to measure the amount of hybridization at each position on the microarray with an Affymetrix scanner (Affymetrix, Santa Clara, CA). For each stimulus a time series of nucleic acid levels (C={C1,C2,C3,...Cn}) and a corresponding time series of nucleic acid levels (M={M1,M2,M3,...Mn}) in control medium in the same experiment as the stimulus is obtained. Quantitative data is then analyzed. Hybridization analysis using microarray is only one method for obtaining gene expression values. Other methods for obtaining gene expression values known in the art or developed in the future can be used with the present invention. Once the gene expression values are determined, the sample can be classified.

Once the gene expression levels of the sample are obtained, the levels are compared or evaluated against a model or control sample(s), and then the sample is

classified, for example, based one whether a particular gene in the sample exhibits increased or decreased expression or whether a marker gene expression pattern is similar to the gene expression pattern specific to a tumor. The evaluation of the sample determines whether or not the sample is assigned to a particular tumor class, or whether or not a candidate compounds modulates tumor development.

By "a marker gene expression pattern similar to the gene expression pattern specific to a tumor" is meant that a marker gene is expressed at least 50%, more preferably, at least 60%, 70%, 80%, or 90% , and most preferably at least 95% of the level of a gene specific to a tumor, for example those genes described in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3. Such determinations can be made using methods described herein, as well as methods known in the art. Preferably, when more than one marker gene is being assessed in a give sample, each marker gene is expressed at least 50%, more preferably, at least 60%, 70%, 80%, or 90% , and most preferably at least 95% of the level of a gene specific to a tumor.

The correlation between gene expression and classification can be determined using a variety of methods. Methods for defining classes and classifying samples are described, for example, in U.S. Patent Application Serial No. 09/544,627, filed April 6, 2000 by Golub et al., the teachings of which are incorporated herein by reference in their entirety. The information provided by the present invention, alone or in conjunction with other test results, aids in sample classification.

In another embodiment of the invention, a sample is obtained from an individual and an assessment of the expression pattern of a set of marker genes described herein is performed to predict or aid in the prediction or diagnose or aid in the diagnosis of cancer in an individual. A biological sample is obtained from the individual, and the gene expression pattern of a set of genes identified in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS.

13A-13N2, and FIGS. 14A-14A3 is determined. For example, the nucleic acid molecules within the sample can be rendered available for hybridization to an oligonucleotide array as described in the Examples. Alternatively, the expression of the proteins encoded by a set of marker genes identified herein can be assessed, e.g., using antibody-based methods. The marker genes (or encoded proteins) to be assessed can be all or a portion of the marker genes associated with a single particular tumor class, or can be all or a portion of the marker genes associated with several different tumor classes.

The expression pattern obtained can be compared with the expression pattern for one or more classes of tumors as described herein. If the expression pattern is substantially similar to that of a tumor class identified herein, a prediction or diagnosis of cancer is likely. The expression pattern can also be compared with the expression pattern obtained from corresponding normal tissue as a control. Similarly, the expression pattern of these marker genes can also be assessed to monitor the effects of treatment in a manner similar to that used in the monitoring of prostate specific antigen for prostate cancer treatment.

Many of the methods described herein for assessment of gene expression require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal

transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The gene expression of the marker genes identified herein can be analyzed by a variety of methods known in the art, including, but not limited to, gene isolation and sequencing or hybridization of a specific oligonucleotide with amplified gene products. In a preferred embodiment, analysis is performed using chip-based oligonucleotide arrays as described herein and known in the art.

There are a number of genetic markers indicated in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3 for each tumor class. In the methods of the invention it is not necessary that all of the indicated marker genes for any particular class be assessed, although one can assess all marker genes for a particular tumor class or all marker genes for multiple tumor classes. For example, the expression pattern of a subset of these genes can be assessed. In one embodiment, only a single marker gene specific for a particular tumor class is assessed. In another embodiment, multiple marker genes are assessed, each of which is specific for a different tumor class. In a further embodiment, multiple marker genes are assessed, each of which is specific for the same tumor class. For example, it is preferred that at least 2, preferably at least 5, more preferably at least 8, even more preferably at least 20, and even more preferably at least 50 marker genes (or their encoded proteins) are assessed.

The present invention also features methods for identifying compounds that modulate tumor development. Novel compounds identified as described herein are also the subject of the invention. Such methods involve contacting a sample, for example a cell, cell lysate, tissue, or tissue lysate, with a candidate compound, and detecting a decrease in expression of at least one gene specific to a particular tumor or tumor class. A candidate compound that decreases expression of such gene is a compound for use in modulating tumor development. A decrease in an gene specific to a particular tumor or

tumor class may be identified using any of the methods described herein (or any analogous method known in the art). For example, oligonucleotide array systems described herein may be used to determine whether the addition of a test compound to a sample modulates expression of a gene specific to a particular tumor or tumor class in that sample.

By "modulating tumor development" is meant increasing or decreasing the likelihood that a tumor will form or develop in a subject. The modulation in tumor formation may be the result of contacting a sample (for example, a cell, tissue, cell or tissue lysate, nucleic acid, or polypeptide) with a candidate compound. It will be appreciated that the degree of modulation provided by a candidate compound in a given assay will vary, but that one skilled in the art can determine the statistically significant change or a therapeutically effective change in the degree or rate of tumor development.

By "tumor development" is meant the formation or progression of a tumor. As used herein leukemias and lymphomas are considered to be types of tumors. Methods for monitoring tumor development are known to those skilled in the art.

By a "candidate compound" is meant a molecule, be it naturally-occurring or artificially derived, that is surveyed for its effects on the gene expression pattern of a marker gene, employing methods described herein. Examples of candidate compounds include, but are not limited to peptides, polypeptides, synthetic organic molecules, naturally occurring organic molecules, nucleic acid molecules, and combinations thereof.

By "decrease in gene expression" is meant a lowering of the level or expression of, and/or the activity of, one or more genes specific to a particular tumor or tumor class in a cell, tissue, cell lysate, or tissue lysate sample relative to a control sample. A decrease in gene expression may occur, for example, when the sample is contacted with a candidate compound for use in modulating tumor development. The control sample may be a cell, tissue, cell lysate, or tissue lysate that was not contacted with the candidate compound or that was contacted with candidate compound vehicle only. Preferably, the decrease in gene expression of a gene specific to a particular tumor or

tumor class is at least 25%, more preferably, the decrease is at least 50%, 60%, 70%, 80%, or 90% and most preferably, the decrease is at least one-fold, relative to a control sample.

The expression level of an gene specific to a particular tumor or tumor class may be modulated by modulating transcription, translation, or mRNA or protein turnover, or the activity of the gene expression product, and such modulation may be detected using known methods for measuring mRNA and protein levels and activities, e.g., oligonucleotide microarray hybridization, RT-PCR, and ELISA and nucleic acid and protein binding assays.

While the above described candidate compound screening methods are designed primarily to identify candidate compounds that may be used to decrease tumor development, identification of candidate compounds that increases tumor development is also a feature of the present invention. Such candidate compound identification methods involve contacting a sample, for example, a cell, cell lysate, tissue, or tissue lysate with a candidate compound, and detecting an increase in expression of at least one gene specific for a particular tumor or tumor class. A candidate compound that increases expression of such a gene specific to a particular tumor or tumor class is a compound for use in modulating tumor development.

By "increase in gene expression" is meant a raising of the level of expression, and/or the activity, of one or more genes specific to a particular tumor or tumor class in a cell, tissue, cell lysate, or tissue lysate sample relative to a control sample. An increase in gene expression may occur, for example, when the sample is contacted with a candidate compound for use in modulating tumor development. The control sample may be a cell, tissue, cell lysate, or tissue lysate that was not contacted with the candidate compound or that was contacted with candidate compound vehicle only. Preferably, the increase is at least 1.5-fold, more preferably the increase is at least 2-fold, 5-fold, or 10-fold, and most preferably, the increase is at least 20-fold, relative to a control sample.

In general, novel drugs for modulation of tumor development can be identified from large libraries of natural products or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the exemplary methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available, e.g., Chembridge (San Diego, CA). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceangraphics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are generated, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their tumor development-modulatory activities should be employed whenever possible.

When a crude extract is found to modulate (i.e., stimulate (increase) or inhibit (decrease)) tumor development, further fractionation of the positive lead extract is desirable to isolate chemical constituents responsible for the observed effect. Thus, the

goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract having an activity that increases or deceases. The same assays described herein for the detection of activities in mixtures of compounds can be used to purify the active component and to test derivatives thereof. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for treatment are chemically modified according to methods known in the art. Compounds identified as being of therapeutic value may be subsequently analyzed using animal models for diseases, in which it is desirable to increase or decrease tumor development.

The present invention also features arrays, for example, microarrays that have a plurality of oligonucleotide probes involved in tumor development immobilized thereon. The oligonucleotide probe may be specific for one or more genes specific for a particular tumor or tumor class, selected from those genes described herein. Such genes can be obtained using their GenBank Accession Numbers identified in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3. Methods for making oligonucleotide microarrays are well known in the art, and are described, for example, in WO 95/11995, the entire teachings of which are hereby incorporated by reference.

The present invention also provides information regarding the genes that are important in tumor development, thereby providing additional targets for diagnosis and therapy. It is clear that the present invention can be used to generate databases comprising genes specific to a particular tumor or tumor class that will have many applications in medicine, research and industry; such databases are also within the scope of the invention.

The invention will be further illustrated by the following non-limiting examples. The teachings of all references cited herein are incorporated herein by reference in their entirety.

EXAMPLES

Materials and Methods

Approximately 300 human tumor and normal tissue specimens were identified and obtained or purchased from a variety of academic or commercial sources. These specimens represented 30 individual classes of tumor or normal tissue with each class containing between 5 and 20 samples. Total RNA was isolated from these specimens using standard laboratory protocols. "Target" (biotinylated) fragmented complementary RNA (cRNA) was produced from each sample using an established molecular biology protocol. Each Target was hybridized sequentially to two high density Affymetrix oligonucleotide microarrays (Hu6800FL and Hu35KsubA; Affymetrix, Inc., Santa Clara, CA), and gene expression profiles (patterns) were measured using a modified confocal laser scanner according to the manufacturer's instructions.

Analysis of Expression Profile (Pattern) Data

Raw expression data was combined into a master data set containing the expression values for between 6800 and 16,000 genes expressed by each individual sample. A filter was applied to this data set which only allowed those genes expressed at 3-fold above baseline and with an absolute difference in expression value of 100 to pass. A signal-to-noise metric (S2N = mean of class #1-mean of class #2 / standard deviation of class #1 + standard deviation of class #2) was applied to this filtered data set to determine which genes are expressed in each individual class versus the other classes. Finally, by comparing the sets of genes which are expressed specifically in one class of tumor (e.g., pancreatic adenocarcinoma) versus its accompanying normal tissue (e.g., normal pancreas), we have determined sets of genes which are specific to various tumors and their normal tissue counterparts. The results are shown in FIGS. 1A-1R2, FIGS. 2A-2T2, FIGS. 3A-3Z2, FIGS. 4A-4S2, FIGS. 5A-5M2, FIGS. 6A-6W2, FIGS. 7A-7D3, FIGS. 8A-8X2, FIGS. 9A-9C3, FIGS. 10A-10P2, FIGS. 11A-11O2, FIGS. 12A-12V2, FIGS. 13A-13N2, and FIGS. 14A-14A3.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.